Cell Reports, Volume *16*

Supplemental Information

The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease

Yalin Liao, Alfredo Castello, Bernd Fischer, Stefan Leicht, Sophia Föehr, Christian K. Frese, Chikako Ragan, Sebastian Kurscheid, Eloisa Pagler, Hao Yang, Jeroen Krijgsveld, Matthias W. Hentze, and Thomas Preiss

SUPPLEMENTAL FIGURES

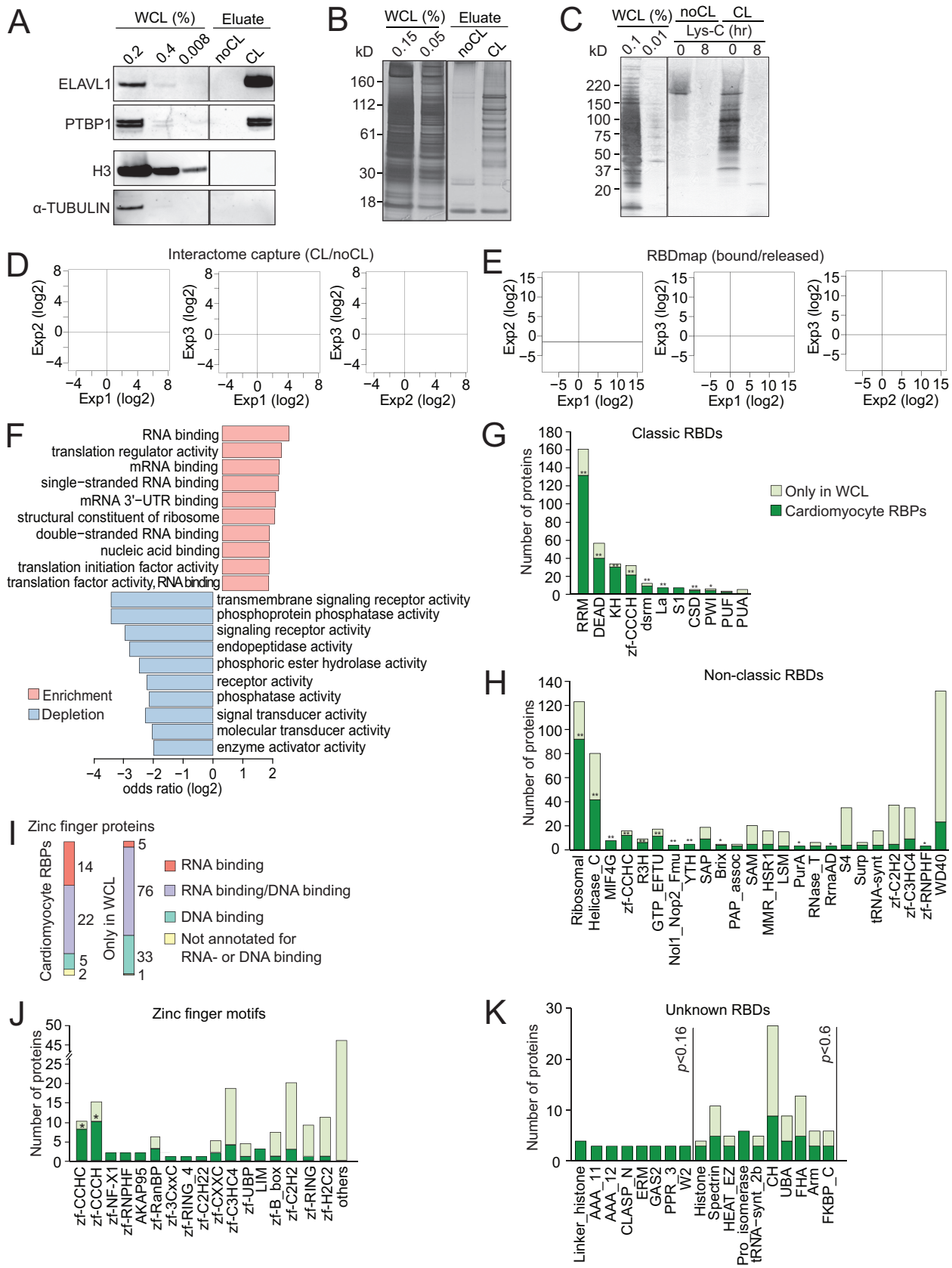

A
WCL (%)
Eluate
0.2
0.4
0.008
noCL
CL
ELAVL1
PTBP1
H3
α-TUBULIN
B
WCL (%)
Eluate
kD
0.15
0.05
noCL
CL
160
112
61
30
18
C
WCL (%)
noCL
CL
Lys-C (hr)
kD
0.1
0.01
0
8
0
8
220
150
100
75
50
37
20
D
Interactome capture (CL/noCL)
Exp2 (log2)
Exp3 (log2)
Exp1 (log2)
Exp2 (log2)
E
RBDmap (bound/released)
F
RNA binding
translation regulator activity
mRNA binding
single−stranded RNA binding
mRNA 3'−UTR binding
structural constituent of ribosome
double−stranded RNA binding
nucleic acid binding
translation initiation factor activity
translation factor activity, RNA binding
transmembrane signaling receptor activity
phosphoprotein phosphatase activity
signaling receptor activity
endopeptidase activity
phosphoric ester hydrolase activity
receptor activity
phosphatase activity
signal transducer activity
molecular transducer activity
enzyme activator activity
Enrichment
Depletion
odds ratio (log2)
G
Classic RBDs
Number of proteins
Only in WCL
Cardiomyocyte RBPs
RRM
DEAD
KH
zf-CCCH
dsrm
La
S1
CSD
PWI
PUF
PUA
H
Non-classic RBDs
Number of proteins
Ribosomal
Helicase_C
MIF4G
zf-CCHC
R3H
GTP_EFTU
Nol1_Nop2_Fmu
YTH
SAP
Brix
PAP_assoc
SAM
MMR_HSR1
LSM
PurA
RNase_T
RrnaAD
S4
Surp
tRNA-synt
zf-C2H2
zf-C3HC4
zf-RNPHF
WD40
I
Zinc finger proteins
Cardiomyocyte RBPs
Only in WCL
14
22
5
2
5
76
33
1
RNA binding
RNA binding/DNA binding
DNA binding
Not annotated for
RNA- or DNA binding
J
Zinc finger motifs
Number of proteins
zf-CCHC
zf-CCCH
zf-NF-X1
zf-RNPHF
AKAP95
zf-RanBP
zf-3CxxC
zf-RING_4
zf-C2H2_2
zf-CXXC
zf-C3HC4
zf-UBP
LIM
zf-B_box
zf-C2H2
zf-RING
zf-H2C2
others
K
Unknown RBDs
Number of proteins
p<0.16
p<0.6
Linker_histone
AAA_11
AAA_12
CLASP_N
ERM
GAS2
PPR_3
W2
Histone
Spectrin
HEAT_EZ
Pro_isomerase
tRNA-synt_2b
CH
UBA
FHA
Arm
FKBP_C

Figure S1

Figure S1. Efficacy and Reproducibility of mRNA Interactome Capture and RBDmap; GO and RBD Enrichment Analyses of Cardiomyocyte RBPs. Related to Figure 1.

(A-B) In the RBDmap approach, RNA-protein complexes were digested with RNases after a first round of oligo(dT) capture, resolved by SDS-PAGE alongside aliquot of input WCL (% equivalent to loaded eluate amount as indicated) and analyzed by western blot (A) (see Supplemental Experimental Procedures for antibody details) or silver stain (B). n=3.

(C) RNA-protein complexes from an RBDmap experiment as in (A) were digested with Lys-C protease, and analyzed by silver stain as in (B).

(D) Protein enrichment in CL over noCL eluates, based on differential LC-MS/MS intensities from three biological replicate mRNA interactome capture experiments, were plotted against each other. Each dot represents one protein group. Protein groups enriched in CL or noCL at 1% FDR are depicted by red or black dots, respectively. Grey dots are protein groups that lack enrichment.

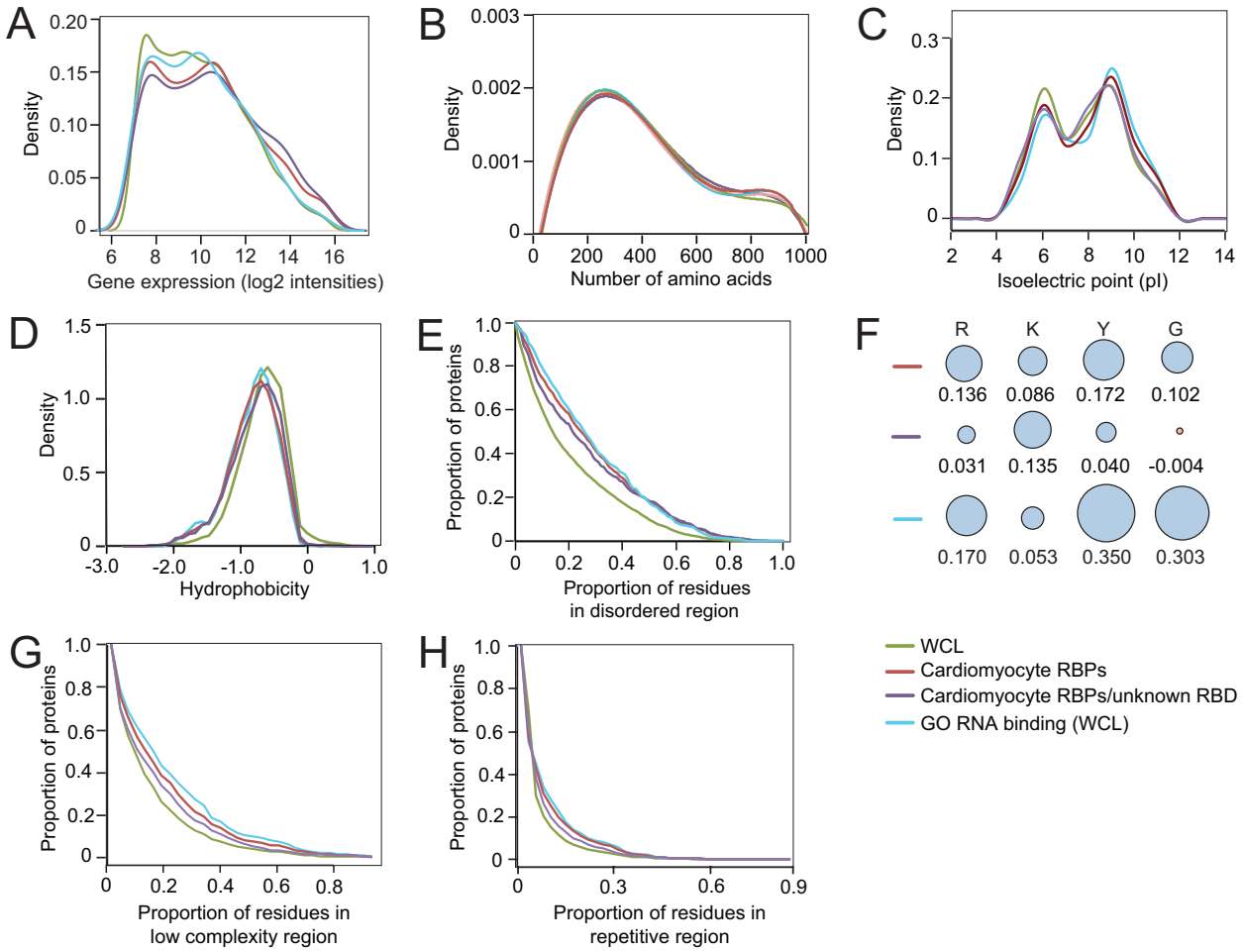
(E) Peptide enrichment in RNA bound fraction over the released fraction, based on differential LC-MS/MS peptide intensity ratio from three biological replicate RBDmap experiments, were plotted against each other. Each dot represents one peptide. Npeps are depicted by red dots. Grey dots indicate peptides that lack enrichment.

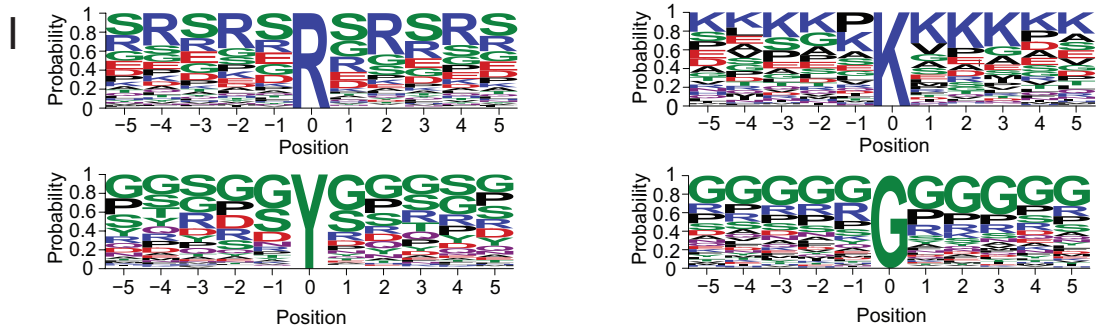
(F) The ten most significantly over- (pink) and under-represented (blue) GOMF terms in cardiomyocyte RBPs, compared to the WCL proteome.

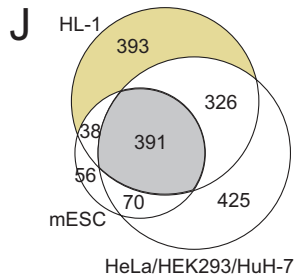
(G) Number of proteins with classic RBDs, in cardiomyocyte RBPs (dark green) or only identified in WCL (light green).

(H) Number of proteins with non-classic RBDs, in cardiomyocyte RBPs (dark green) or only identified in WCL (light green). Only domains represented by three or more hits are shown. Proteins with both classic and non-classic RBDs are considered only in (G).

(I) Distribution of zinc finger-containing proteins with GO ‘RNA binding’, ‘DNA binding’ and their children terms, within cardiomyocyte RBPs (left) or only identified in WCL (right).
(J) Number of proteins with different zinc finger motif subtypes in cardiomyocyte RBPs (dark green) or only identified in WCL (light green).
(K) Number of proteins previously unrelated to RNA biology in cardiomyocyte RBPs (dark green) or only identified in WCL (light green). Only domains represented by three or more hits are shown. Proteins with either classic or non-classic RBDs, are considered only in (G) and (H).


A
Density
Gene expression (log2 intensities)
B
Density
Number of amino acids
C
Density
Isoelectric point (pI)
D
Density
Hydrophobicity
E
Proportion of proteins
Proportion of residues in disordered region
F
R
K
Y
G
0.136
0.086
0.172
0.102
0.031
0.135
0.040
-0.004
0.170
0.053
0.350
0.303
G
Proportion of proteins
Proportion of residues in low complexity region
H
Proportion of proteins
Proportion of residues in repetitive region
WCL
Cardiomyocyte RBPs
Cardiomyocyte RBPs/unknown RBD
GO RNA binding (WCL)


I
Probability
Position


J
HL-1
393
326
38
391
56
70
425
mESC
HeLa/HEK293/HuH-7

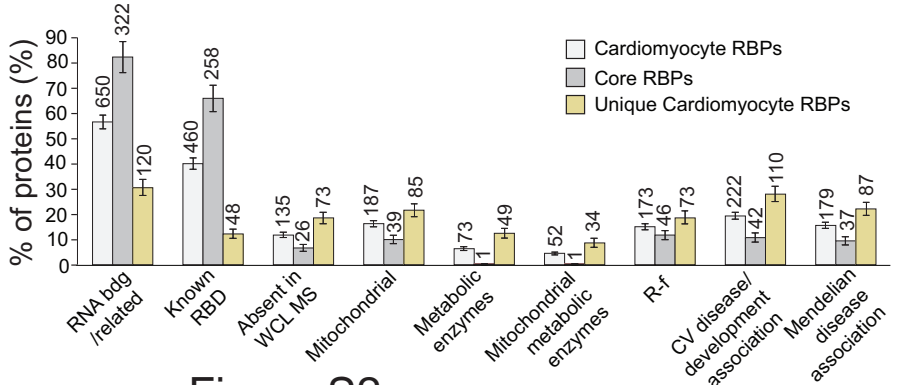

K
% of proteins (%)
Cardiomyocyte RBPs
Core RBPs
Unique Cardiomyocyte RBPs
650
322
120
460
258
48
135
26
73
187
39
85
73
1
49
52
1
34
173
46
73
222
42
110
179
37
87
RNA bdg /related
Known RBD
Absent in WCL MS
Mitochondrial
Metabolic enzymes
Mitochondrial metabolic enzymes
R-f
CV disease/ development association
Mendelian disease association

Figure S2

Figure S2. Biophysical and Sequence Features of Cardiomyocyte RBPs. Related to Figure 1.

The following proteins sets were used in panels A-H: WCL proteins (green; used as background in panel F), cardiomyocyte RBPs (red), cardiomyocyte RBPs with unknown RBDs (purple). WCL RBPs with GO 'RNA binding' and its children terms are shown for a reference (blue). For information on bioinformatic methods see Supplemental Experimental Procedures and (Castello et al., 2012; Castello et al., 2013b).

(A) Density of mRNA expression levels in normal HL-1 cells (GEO: GSE56584).

(B) Density of calculated protein length.

(C) Density of calculated isoelectric point (pI).

(D) Density of hydrophobicity.

(E) Proportion of amino acid residues in disordered regions.

(F) Amino acids composition in disordered region of all cardiomyocyte RBPs, cardiomyocyte RBPs with unknown RBDs, or all WCL proteins with GO 'RNA binding' and its children terms. Blue circles indicate enrichment and pink circles indicate depletion relative to the composition of disordered regions in all WCL proteins. Circle area is proportional to the log2 enrichment or depletion ratio.

(G) Distribution of calculated low complexity regions.

(H) Distribution of repetitive dipeptide sequence (dimers of subsequent amino acids, with 0, 1, or 2 amino acids spacing).

The significance of differences between RBP subsets in panels A, B, C, D, E, G, H was tested by two-sample Kolmogorov-Smirnov tests. This showed that distribution of protein size (Number of amino acids) does not differ between all four groups. Compared to WCL, cardiomyocyte RBPs and cardiomyocyte RBPs with unknown RBDs are significantly different in mRNA expression

levels, all three subsets are significantly different in, isoelectric point, hydrophobicity, disordered region, low-complexity region, and repetitive region ($p < 0.01$).

The significance of enrichment/depletion in panel F was tested by two-sample test for population proportions. Comparing all other groups to the WCL proteome, results were significant ($p < 0.01$) for all amino acids, except for glycine (G) in the cardiomyocyte RBP/unknown RBD group.

(I) 11-mers around the 4 repetitive amino acid residues featured in (F) were extracted and aligned by their central residue to show sequence bias. Height of letter indicates the probability of occurrence of a given amino acid at each position.

(J) Venn diagram comparing cardiomyocyte RBPs with four published mRNA interactome datasets: HeLa (Castello et al., 2012), HEK293 (Baltz et al., 2012), mESC (Kwon et al., 2013), and HuH-7 (Beckmann et al., 2015).

(K) Distribution of a range of features among all cardiomyocyte RBPs, core RBPs and unique cardiomyocyte RBPs. Percentage uncertainties were derived using Gaussian error propagation assuming Poisson uncertainties of protein counts. Cardiomyocyte RBPs, 1148; Core RBPs, 391; Unique Cardiomyocyte RBPs, 393.

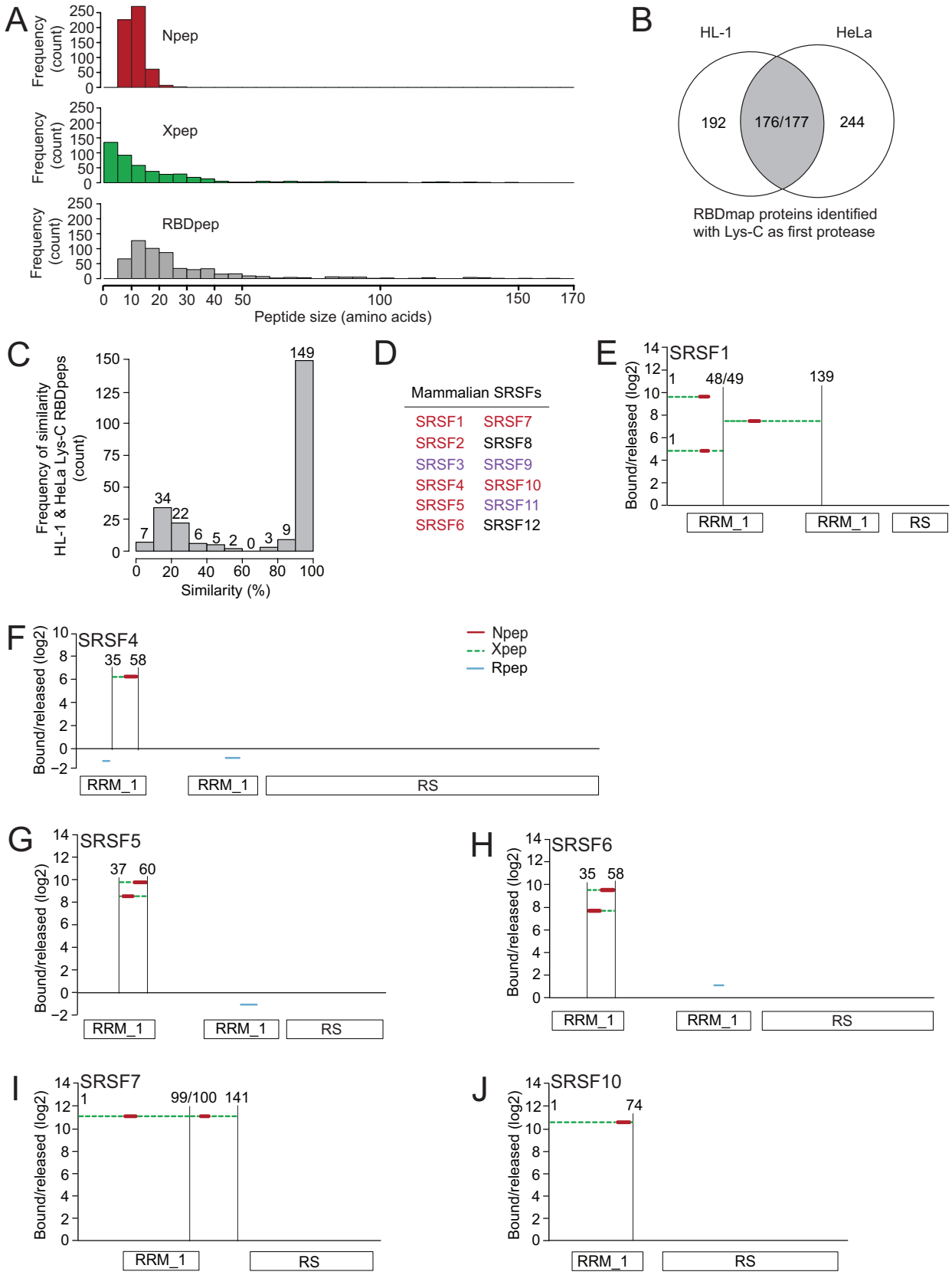

A
Npep
Xpep
RBDpep
Frequency (count)
Peptide size (amino acids)
B
HL-1
HeLa
192
176/177
244
RBDmap proteins identified with Lys-C as first protease
C
Frequency of similarity HL-1 & HeLa Lys-C RBDpeps (count)
Similarity (%)
D
Mammalian SRSFs
SRSF1
SRSF2
SRSF3
SRSF4
SRSF5
SRSF6
SRSF7
SRSF8
SRSF9
SRSF10
SRSF11
SRSF12
E
SRSF1
F
SRSF4
Npep
Xpep
Rpep
G
SRSF5
H
SRSF6
I
SRSF7
J
SRSF10
Bound/released (log2)
RRM_1
RS

Figure S3

Figure S3. Benchmarking HL-1 RBDmap data. Related to Figure 2.

(A) Distribution of Npeps, Xpeps (≥ 1 amino acid) and RBDpeps identified in HL-1 RBDmap.

(B) Venn diagram comparing homologous proteins in HL-1 & HeLa RBDmap data (Castello et al., 2016). Both ENSG00000092199 and ENSG00000179172 are human homologs of mouse ENSMUSG00000060373.

(C) Distribution of 237 HL-1 RBDpeps identified from 176 proteins that are also in HeLa RBDmap (Castello et al., 2016), based on sequence similarity (See Supplemental Experimental Procedures).

(D) List of SRSF splicing factor family members found in the cardiomyocyte RBPs (purple) and also in RBDmap (red).

(E-J) RBDmap data for six SRSF family members, SRSF1 (E), SRSF4 (F), SRSF5 (G), SRSF6 (H), SRSF7 (I) and SRSF10 (J), shown as in Figure 2E. See Supplemental Experimental Procedure for RS domain annotation.

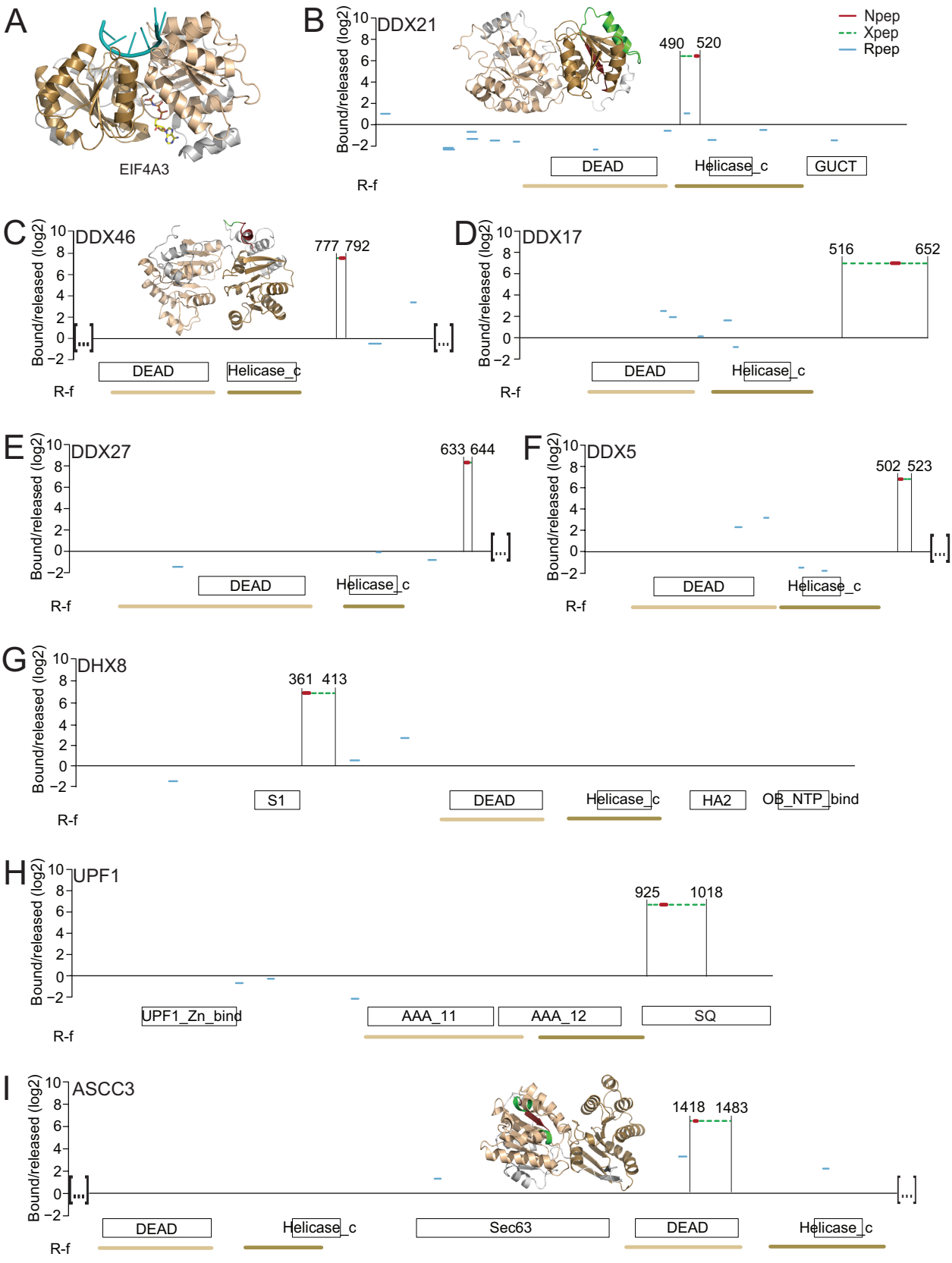

A
EIF4A3
B
DDX21
490 520
Npep
Xpep
Rpep
Bound/released (log2)
DEAD
Helicase_c
GUCT
R-f
C
DDX46
777 792
DEAD
Helicase_c
D
DDX17
516
652
E
DDX27
633 644
F
DDX5
502 523
G
DHX8
361 413
S1
HA2
OB_NTP_bind
H
UPF1
925 1018
UPF1_Zn_bind
AAA_11
AAA_12
SQ
I
ASCC3
1418 1483
Sec63

Figure S4

Figure S4. RNA-Binding Modalities in Seven RNA Helicases and a DNA Helicase. Related to Figure 4.

(A) Crystal structure of EIF4A3 (EC:3.6.4.13), shown as a ribbon diagram, in complex with 5'-UUUUUU-3' RNA (teal) and adenosine-5'-triphosphate (ATP, multicolor) (PDB: 2HYI, chain C). N-terminal and C- terminal Rossmann folds (R-f) are depicted in wheat and sand, respectively.

(B-H) RBDmap data for seven RNA helicases, DDX21 (B), DDX46 (C), DDX17 (D), DDX27 (E), DDX5 (F), DHX8 (G) and UPF1 (H) (EC:3.6.4.13), as shown in Figure 2E. RBDpeps were mapped to Phyre2-modeled structures of DDX21, DDX46, DDX5 (not shown) and DDX27 (not shown). Rossmann folds are depicted in wheat or sand.

(I) RBDmap data for ASCC3 (EC:3.6.4.12). RBDpep was mapped to Phyre2-modeled structure.

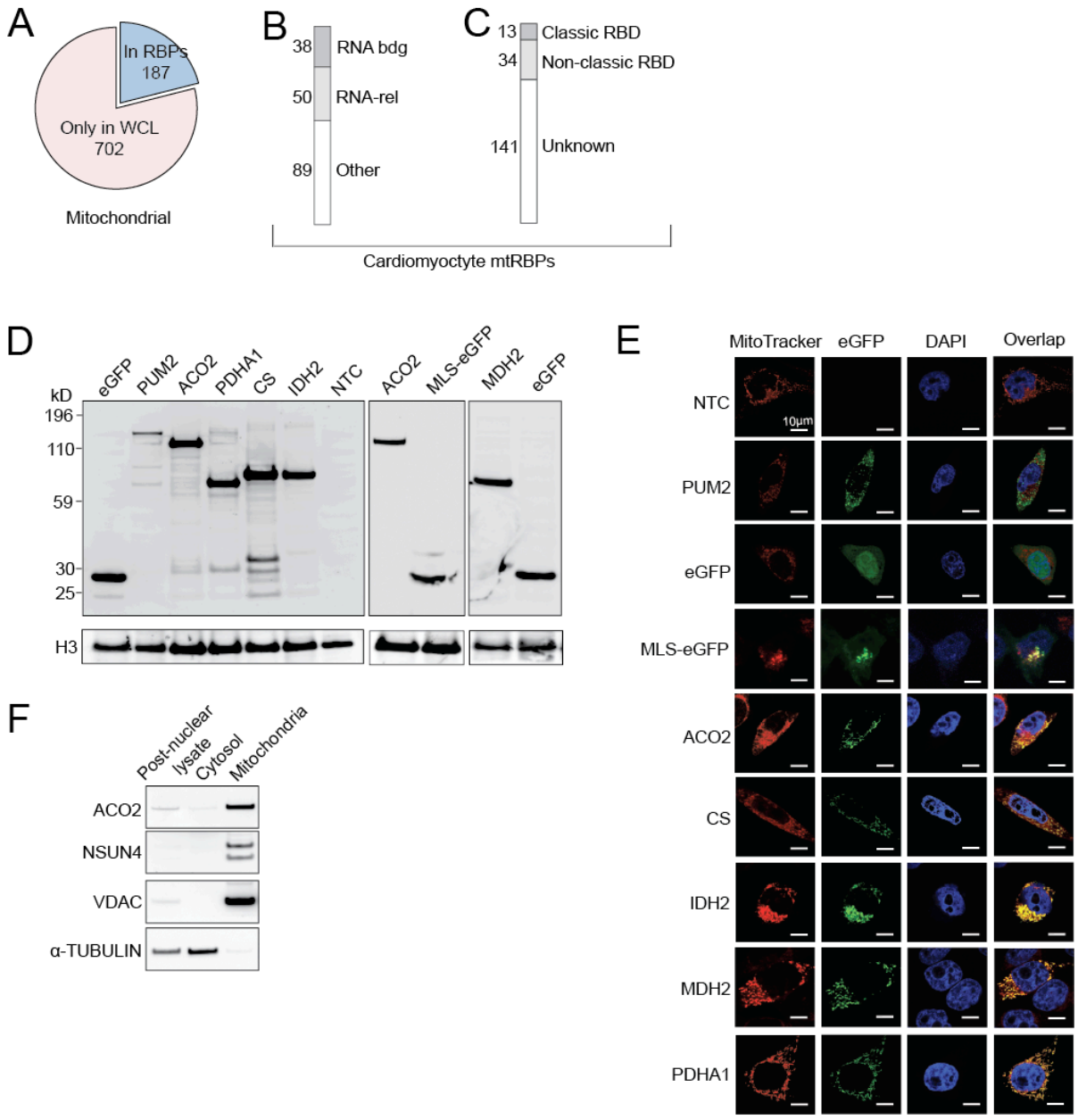

A
In RBPs
187
Only in WCL
702
Mitochondrial
B
38
RNA bdg
50
RNA-rel
89
Other
C
13
Classic RBD
34
Non-classic RBD
141
Unknown
Cardiomyoctyte mtRBPs
D
eGFP
PUM2
ACO2
PDHA1
CS
IDH2
NTC
ACO2
MLS-eGFP
MDH2
eGFP
kD
196
110
59
30
25
H3
E
MitoTracker
eGFP
DAPI
Overlap
NTC
10μm
PUM2
eGFP
MLS-eGFP
ACO2
CS
IDH2
MDH2
PDHA1
F
Post-nuclear lysate
Cytosol
Mitochondria
ACO2
NSUN4
VDAC
α-TUBULIN

Figure S5

Figure S5. Characterization of Cardiomyocyte mtRBPs. Related to Figure 6.

(A) Number of WCL mitochondrial proteins present/absent in cardiomyocyte RBPs.

(B) Proportion of cardiomyocyte mtRBPs with GO 'RNA binding' or RNA-related annotation, shown as in Figure 1E.

(C) Proportion of cardiomyocyte mtRBPs harboring classic, non-classic, or unknown RBDs, shown as in Figure 1F.

(D) Analysis of eGFP fusion protein expression by western blot. Probing was done with anti-GFP antibody (top) or anti-histone H3 as a loading control (bottom). Dose of plasmid DNA based on cell culture area is the same as used in RNA ^{32}P 5' end-labeling experiment. NTC, non-transfected control.

(E) Immunofluorescence analysis of mitochondrial localization for C-terminal tagged eGFP fusion proteins (green) in HeLa cells. Red, MitoTracker; blue, DAPI. Scale bar: 10 μm.

(F) Quality control of HeLa cell mitochondria isolation by western blot, using antibodies against mitochondrial matrix markers (ACO2, NSUN4), mitochondrial outer membrane marker (VDAC) and cytosolic marker (α-TUBULIN).

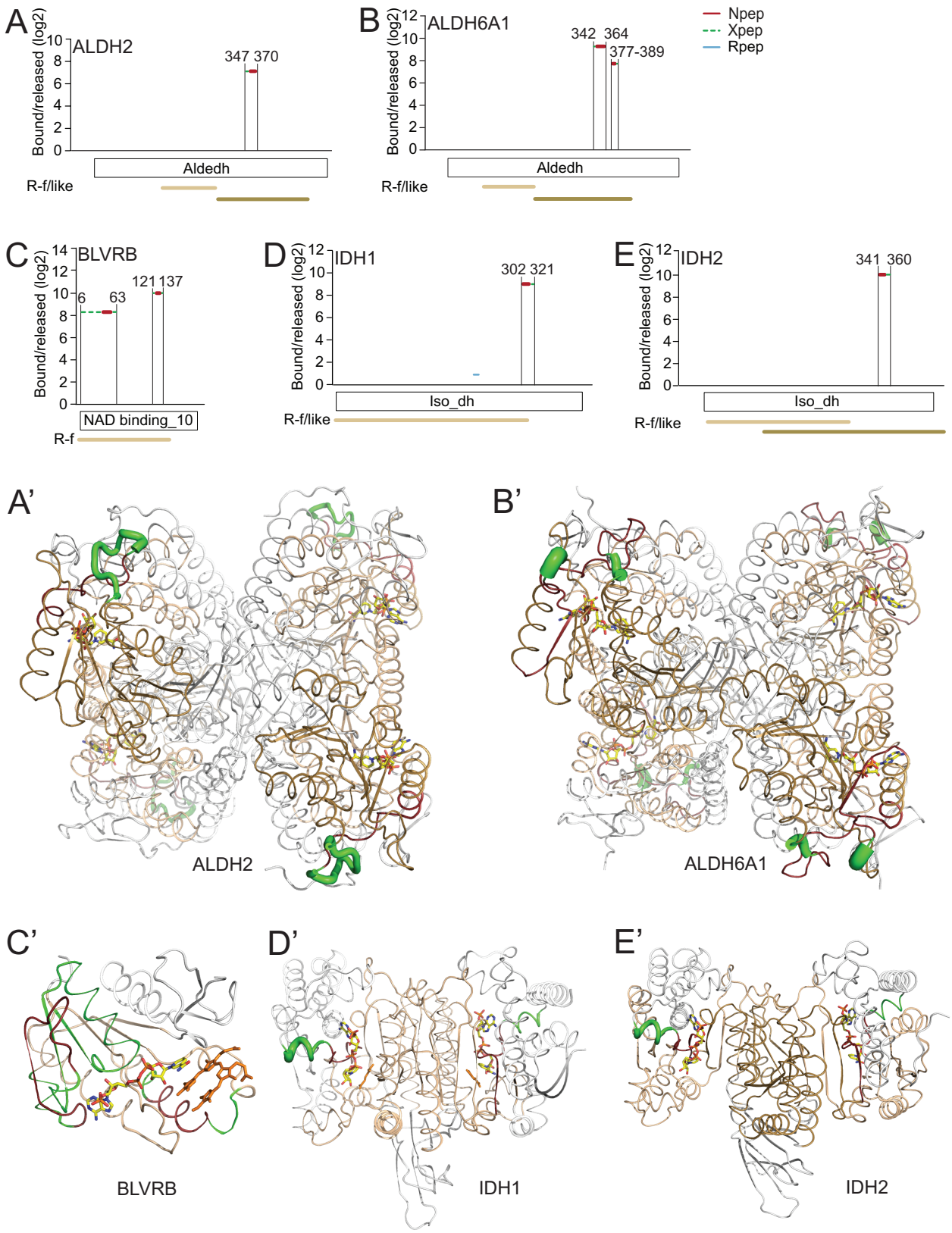

A
ALDH2
Bound/released (log2)
347 370
Aldedh
R-f/like
B
ALDH6A1
342 364
377-389
Aldedh
R-f/like
Npep
Xpep
Rpep
C
BLVRB
6
63
121 137
NAD binding_10
R-f
D
IDH1
302 321
Iso_dh
R-f/like
E
IDH2
341 360
Iso_dh
R-f/like
A'
ALDH2
B'
ALDH6A1
C'
BLVRB
D'
IDH1
E'
IDH2

Figure S6

Figure S6. RNA-Binding Modalities Among Five Rossmann Fold or Rossmann Fold-Like Enzymes. Related to Figure 4, Figure 7.

(A-E) RBDmap data for ALDH2 (EC:1.2.1.3), ALDH6A1 (EC:1.2.1.18, EC:1.2.1.27), BLVRB (EC:1.5.1.30), IDH1 (EC:1.1.1.42) and IDH2 (EC:1.1.1.41), shown as in Figure 2E.

(A'-E') RBDpeps were mapped to enzyme crystal structures.

(A') Crystal structure of ALDH2 tetramer, in complex with NAD^+ (multicolor) (PDB: 1O01).

(B') Crystal structure of ALDH6A1 tetramer, in complex with NAD^+ (multicolor) (PDB: 1T90).

(C') Crystal structure of BLVRB monomer, in complex with biliverdine IX alpha (BLA, orange) and $NADP^+$ Nicotinamide-adenine-dinucleotide phosphate (multicolor) (PDB: 1HE2).

(D') Crystal structure of IDH1 dimer, in complex with 2-oxoglutaric acid (orange) and NADPH (multicolor) (PDB: 3INM).

(E') Crystal structure of IDH2 dimer, in complex with NADPH (multicolor) (PDB: 4JA8).

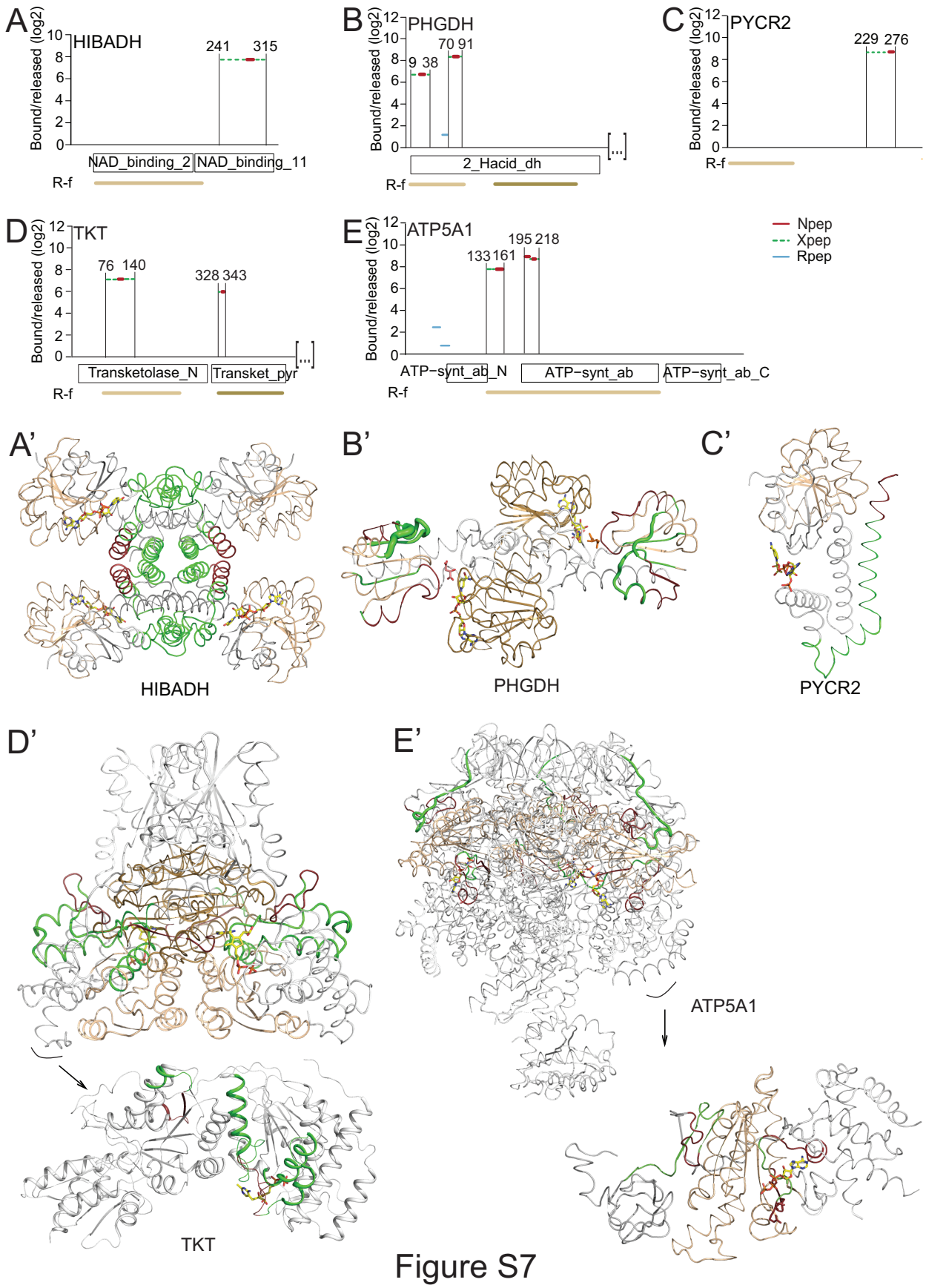

A
HIBADH
241
315
Bound/released (log2)
NAD_binding_2
NAD_binding_11
R-f
B
PHGDH
9
38
70
91
2_Hacid_dh
C
PYCR2
229
276
D
TKT
76
140
328
343
Transketolase_N
Transket_pyr
E
ATP5A1
133
161
195
218
ATP−synt_ab_N
ATP−synt_ab
ATP−synt_ab_C
Npep
Xpep
Rpep
A'
HIBADH
B'
PHGDH
C'
PYCR2
D'
TKT
E'
ATP5A1

Figure S7

Figure S7. RNA-Bind Modalities Among Five Rossmann Fold or Rossmann Fold Like Enzymes. Related to Figure 4, Figure 7.

(A-E) RBDmap data for HIBADH (EC:1.1.1.31), PHGDH (EC:1.1.1.95) PYCR2 (EC:1.5.1.2), TKT (EC:2.2.1.1) and ATP5A1, shown as in Figure 2E.

(A'-E') RBDpeps were mapped to enzyme crystal structures.

(A') Crystal structure of HIBADH tetramer, in complex with NAD^+ (multicolor) (PDB: 2I9P).

(B') Crystal structure of PHGDH dimer, in complex with D-malate (orange) and NAD^+ (multicolor) (PDB: 2G76).

(C') Crystal structure of PYCR2 monomer, in complex with $NADP^+$ (multicolor) (PDB: 2GRA).

(D') Crystal structure of TKT dimer, in complex with D-xylitol-5-phosphate (orange) and thiamin diphosphate (multicolor) (PDB: 4KXV). Arrow indicates enlarged view of monomer crystal structure.

(E') Crystal structure of ATP synthase complex (α3β3γ1δ1ε1) in complex with ATP (multicolor) (PDB: 4YXW). Arrow indicates enlarged view of α subunit ATP5A1.

SUPPLEMENTAL TABLE LEGENDS

Table S1. Mass Spectrometry data of cardiomyocyte WCL proteome and mRNA interactome. Related to Figure 1.

Results are shown as protein groups, their majority Uniprot IDs, and the Ensembl gene IDs. For mRNA interactome capture, the log2 enrichment (CL/noCL) of each protein group in biological replicate samples and statistical significance is also shown. *, detected in whole cell lysate (WCL); #, present in at least two biological replicates.

Table S2. Peptides identified by cardiomyocyte RBDmap. Related to Figures 2, 3, 4, 6, 7.

A total of 3342 tryptic peptides were identified. 771 peptides do not have quantitative RNA bound/released enrichment value and therefore were removed. 453 of the remaining 2571 peptides either do not map to unique genes, or do not represent the majority protein coded by a unique gene, or no gene identifier is available; therefore these were also removed (Table S2. "Removed peptides").

The remaining 2118 peptides were classified into 4 groups:

568 peptides were assigned as Npeps (Figure 1A), covering 368 unique genes. They are enriched in the RNA bound compared to the released fraction (log2 enrichment greater than 0) with an FDR of less than 1%. For each gene, the protein with the highest coverage was selected (Table S2. "Npep").

124 peptides were assigned as Candidate Npeps, they are enriched in the RNA bound compared to the released fraction (log2 enrichment greater than 0) with an FDR between 1% and 10% (Table S2. "Candidate Npep").

1287 peptides were assigned as Rpeps (Figure 1), they are peptides with an FDR greater than 15% regardless of enrichment value (Table S2. “Rpep”).

139 peptides remained unassigned as they did not satisfy the selection criteria stated above (Table S2. “Unassigned peptides”).

Table S3. A collection of cardiomyocyte RBPs features. Related to Figures 2, 3, 4, 6, 7.

Cardiomyocyte RBPs are listed by Ensembl gene ID and gene name, and the presence of the following features are indicated by a ‘+’ for each entry: identified by mRNA interactome capture; identified by RBDmap; mitochondrial localization; metabolic enzyme; Mendelian disease association; Mendelian RBDpep; categorized as a PPIase; associated with cardiovascular disease & development; unique cardiomyocyte RBP; core RBP. The following features are also listed: RNA-related/unrelated annotation; category of RBD; name of known RBD; top ten enriched and depleted GOMF terms listed against the respective RBP; EC number for metabolic enzyme; type of Rossmann fold homologous superfamily; RNA helicase family; type of RNA modification for RNA modification enzymes. * see Supplemental Experimental Procedure for detail; # RBDpep covers Mendelian disease missense mutation/amino acid deletion; § compared to the following mRNA interactome datasets: HeLa (Castello et al., 2012), HEK293 (Baltz et al., 2012), mESC (Kwon et al., 2013), and HuH-7 (Beckmann et al., 2015).

Table S4. Spectrum of OMIM diseases associated with cardiomyocyte RBPs. Related to Figure 3.

Cardiomyocyte OMIM-RBPs are listed by Ensembl gene ID and gene name. For each RBP entry, the associated Mendelian disease is shown by phenotype MIM number, name of disease and type

of disease. For RBPs where the RBDpep covers disease mutation(s), the missense mutation(s) and/or amino acid deletion(s) are also indicated.

Table S5. Characteristics of metabolic enzymes among cardiomyocyte RBPs. Related to Figures 4, 7.

Metabolic enzymes among the cardiomyocyte RBPs are listed by Ensembl gene ID and gene name. The presence of the following features are indicated by a '+' for each entry: Rossmann fold; Rossmann-like fold; mitochondrial localization; presence in the mitochondrial RNA processing granule. The following features are also listed: EC number; EC class; type of Rossmann fold homologous superfamily; metabolic pathway; non-substrate ligand. *, see Supplemental Experimental Procedures for details; #, non-substrate ligand annotation was obtained from Uniprot; § as determined in (Antonicka and Shoubridge, 2015).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell Culture

HL-1 cardiomyocytes (a gift from Dr. William Claycomb, Louisiana State University) were maintained as described (Claycomb et al., 1998; Humphreys et al., 2012). HeLa cells were purchased from ATCC and maintained as described (Clancy et al., 2011). All cells were grown in an atmosphere of 5% CO_2 at 37 °C and the cell culture media were changed every other day.

mRNA Interactome Capture

The procedure was performed as previously described (Castello et al., 2012; Castello et al., 2013b) with some modifications to accommodate the properties of cardiomyocyte cultures. Eighteen to forty-eight 145 cm^2 culture dishes were seeded with HL-1 cells at 6×10^6 cells per dish, and grown for 4 days until confluent to obtain spontaneously beating cells. For in-cell crosslinking (CL), culture dishes were washed twice with ice-cold 1×PBS, placed on ice and irradiated in a Stratalinker (Stratagene) with 150 mJ/cm^2 of UV light at 254 nm. Non-crosslinked cells (noCL) were processed in parallel as a specificity control. Unless otherwise stated, the following gives reagent amounts per 145 cm^2 culture dish. Cells (CL or noCL) were immediately collected in ice-cold 1×PBS, pelleted and lysed in 10 ml lysis buffer (20 mM Tris•HCl pH 7.5, 500 mM LiCl, 0.5% LiDS, 1 mM EDTA, and 5 mM DTT), then incubated on ice for 10 min. The lysate was then passed three times through a narrow gauge needle (0.4 mm diameter), before addition of 0.67 ng polyadenylated spike-in RNA, which was *in vitro* transcribed from plasmid pT3lucpA (Iizuka et al., 1994; Preiss and Hentze, 1998). RNA-protein complexes were captured on oligo$(dT)_{25}$ magnetic bead (beads from 150 µl original bead suspension, New England Biolabs), by incubating for 1 hr at 4 °C on a rotating wheel. Subsequently, beads were collected

on a magnet and washed twice, with 3 ml of each the following buffers (I: 20 mM Tris•HCl pH 7.5, 500 mM LiCl, 0.1% LiDS, 1 mM EDTA, 5 mM DTT; II: 20 mM Tris•HCl pH 7.5, 500 mM LiCl, 1 mM EDTA, 5 mM DTT; III: 20 mM Tris•HCl pH 7.5, 200 mM LiCl, 1 mM EDTA, 5 mM DTT), for 5 min at 4 °C on a rotating wheel. Finally, RNA-protein complexes were eluted by incubation of the beads with 67 µl 20 mM Tris•HCl pH 7.5, 1 mM EDTA at 50 °C for 3 min. After elution, the oligo$(dT)_{25}$ beads were reactivated in lysis buffer according to the manufacturer recommendations, and two additional bead binding/elution rounds were carried out for each sample.

Finally, all three eluates (from the three capture rounds) were combined and an aliquot was taken for qRT-PCR (equivalent to 50 cm^2 culture area), this aliquot was supplemented with ¼ volume of 5×proteinase K buffer (50 mM Tris•HCl pH 7.5, 750 mM NaCl, 1% SDS, 50 mM EDTA, 2.5 mM DTT, 25 mM $CaCl_2$), and treated with proteinase K (Thermo Scientific) at 42 °C for 1 hr, at final concentration of 1 µg/µl, prior to RNA extraction and cDNA synthesis. The remaining eluate was supplemented with ¼ volume of 5×RNase buffer (50 mM Tris•HCl pH 7.5, 750 mM NaCl, 0.25% NP-40, 2.5 mM DTT) and treated with 0.61 U of RNase T1 (Sigma) and 0.2 µg RNase A (Sigma) for 1 hr at 37 °C, further aliquots were taken for protein analysis (equivalent to 100 cm^2 or 200 cm^2 culture area for western blot and silver stain, respectively). The bulk of the eluate was then processed for mass spectrometry.

RBDmap

Initial purification of RBPs was carried out following the RNA interactome capture protocol described above using ten culture dishes per sample. A noCL control was carried out for RNA and protein recovery control purposes. Eluted RNA-protein complexes (1.6 ml/sample) were supplemented with 1 µl (40 U) of RNaseOUT (Life Technologies) and then treated with 1 µg of

endoproteinase Lys-C (Sigma) at 37 °C for 8 hr. An aliquot (75 µl) from each sample was taken prior to and after Lys-C digestion to monitor protein and RNA integrity. The rest of the Lys-C digested sample (1.45 ml) was then diluted with 2.4 ml H_2O and 1 ml 5×dilution buffer (2.5 M LiCl, 100 mM Tris•HCl pH 7.5, 5 mM EDTA, 25 mM DTT), and incubated with oligo$(dT)_{25}$ magnetic beads (beads from 2 ml original bead suspension, New England Biolabs) for 1 hr at 4 °C on a rotating wheel. Beads were collected with a magnet, and the supernatant was kept, representing the released fraction. Beads were washed once with lysis buffer, buffer I, II and III before elution of the RNA bound fraction from the beads following the elution protocol described above. Three bead binding/elution rounds were carried out as described for mRNA interactome capture above. Both, released and RNA bound fractions were treated with RNase T1 and RNase A and processed for mass spectrometry. Details on the RBDmap method development are described elsewhere (Castello et al., 2016).

HL-1 Whole Cell Lysate

2.3×10^6 HL-1 cells were seeded on 56.7 cm^2 culture dish, and grown for 4 days. Cells were washed twice with ice-cold 1×PBS, and immediately collected in 5 ml ice-cold 1×PBS, pelleted and incubated on ice in 4 ml lysis buffer (as described in Supplemental Experiment Procedure mRNA interactome capture section). The lysate was then passed three times through a narrow gauge needle (0.4 mm diameter), incubated on ice for 10 min, followed by Trichloroacetic acid (TCA) precipitation: 1 ml of ice-cold TCA was added to the lysate, then incubated on ice for 30 min. The lysate was centrifuged at 14,000 × *g* for 20 min. The supernatant was removed, the pellet was then washed with 1 ml of 10% TCA (in ddH_2O) and vortexed before centrifugation at 14,000 × *g* for 20 min. The wash step was repeated with 1 ml of cold acetone (chilled at -20 °C).

The pellet was air-dried, then processed for mass spectrometry. All centrifugations were performed at 4 °C.

Mass Spectrometry

Sample Preparation for Mass Spectrometry

Samples were processed according to the standard filter aided sample preparation protocol with minor modifications (Wisniewski et al., 2009). Cysteines were reduced (10 mM DTT, 55 °C, 30 min) and alkylated (20 mM Iodoacetamide, 30 min in the dark).

HL-1 cell lysate and mRNA interactome capture samples were processed according to the following protocol: Samples were buffer-exchanged into digestion buffer (8 M Urea, 50 mM Triethylammoniumbicarbonate (TEAB, Sigma-Aldrich)) using 10 kD centrifugal devices (Millipore) and incubated with 1 µg sequencing grade Lys-C/Trypsin (Promega) at 37 °C for 4 hr. The buffer was diluted with 50 mM TEAB to a final Urea concentration of <2 M and incubated over night at 37 °C. Resulting peptides were desalted and labeled using stable isotope reductive methylation (Boersema et al., 2009) on OASIS HLB solid phase extraction cartridges (Waters). Labels were swapped between replicates. Peptides were fractionated into 12 fractions on a 3100 OFFGEL Fractionator (Agilent) using Immobiline DryStrips (pH 3-10 NL, 13 cm; GE Healthcare) according to the manufacturer's protocol. Isoelectric focusing was carried out at a constant current of 50 mA allowing a maximum voltage of 8000 V. When 20 kVh were reached the fractionation was stopped, fractions were collected and desalted using StageTips. Samples were dried in a vacuum concentrator and reconstituted in MS loading buffer (5% DMSO, 1% formic acid).

RBDmap samples were processed according to the following protocol: RNA bound and RNA released fractions were buffer-exchanged into 50 mM TEAB using 3 kD centrifugal devices

(Millipore) and incubated with 1 µg sequencing grade Trypsin (Promega) at 37 °C over night. Resulting peptides were desalted and labeled using stable isotope reductive methylation on OASIS HLB solid phase extraction cartridges. Labels were swapped between replicates. Peptides were fractionated by offline high pH reversed phase chromatography (Yang et al., 2012). Briefly, peptides were resuspended in buffer A (20 mM Ammoniumformate, pH 10) and fractionated over a Gemini 3U C18 110A column (100 × 1.00 mm, Phenomenex) using a 60 min linear gradient from 0-35% solvent B (100% Acetonitrile) at a constant flow rate of 0.1 mL/min. Resulting fractions were pooled into 8 samples, desalted, dried in a vacuum concentrator and reconstituted in MS loading buffer.

LC-MS/MS

Samples were analyzed on an LTQ Orbitrap Velos Pro mass spectrometer (Thermo Scientific) coupled to a nanoAcquity UPLC system (Waters). Peptides were loaded onto a trapping column (nanoAcquity Symmetry C_{18}, 5 µm, 180 µm × 20 mm) at a flow rate of 15 µl/min with solvent A (0.1% formic acid). Peptides were separated over an analytical column (nanoAcquity BEH C18, 1.7 µm, 75 µm × 200 mm) at a constant flow of 0.3 µl/min using the following gradient: 3% solvent B (Acetonitrile, 0.1% formic acid) for 10 min, 7-25% solvent B within 160 min, 25-40% solvent B within 10 min, 85% solvent B for 6 min. Peptides were introduced into the mass spectrometer using a Pico-Tip Emitter (360 µm outer diameter × 20 µm inner diameter, 10 µm tip, New Objective). MS survey scans were acquired from 300-1700 m/z at a nominal resolution of 30,000. The 15 most abundant peptides were isolated within a 2 Da window and subjected to MS/MS sequencing using collision-induced dissociation in the ion trap (activation time 10 msec, normalized collision energy 40%). Only 2+/3+ charged ions were included for analysis. Precursors were dynamically excluded for 30 sec (exclusion list size was set to 500).

Peptide Identification and Quantification

Raw data were processed using MaxQuant (version 1.3.0.5) (Cox and Mann, 2008). MS/MS spectra were searched against the UniProt mouse database (WCL and interactome capture: version 12_2013, RBDmap: version 03_2014) concatenated to a database containing protein sequences of common contaminants. Enzyme specificity was set to trypsin/P, allowing a maximum of two missed cleavages. Cysteine carbamidomethylation was set as fixed modification, and methionine oxidation and protein N-terminal acetylation were used as variable modifications. The minimal peptide length was set to six amino acids. The mass tolerances were set to 20 ppm for the first search, 6 ppm for the main search and 0.5 Da for product ion masses. False discovery rates (FDR) for peptide and protein identification were set to 1%. Match between runs (time window 2 min) and re-quantify options were enabled.

Definition of mRNA interactome proteins

Statistical analysis for CL/noCL enrichment of protein groups quantified in at least two out of three biological replicates was performed using an empirical Bayes moderated t-test within the R/Bioconductor package limma (Smyth, 2004). p-values were adjusted for multiple testing using the method of Benjamini-Hochberg by controlling for FDR. The UniProt accession numbers of each protein group passing the criteria log2 enrichment >0 at an FDR of 1% were converted into Ensembl gene IDs (Ensembl release 72). Where multiple Ensembl gene IDs applied, the lowest Ensembl gene ID was used. The duplicate Ensembl gene IDs were then removed, and majority proteins corresponding to the remaining genes were defined as the mRNA interactome.

Definition of RBDpeps

To identify RNA-binding sites, the log2 intensity ratio of MS-identified peptides in the RNA bound to the released fraction was considered. The distribution of the log2 ratios is bi-modal, representing the released and RNA bound peptides. The log2 ratios are normalized to the location of the left mode using a robust estimate. Log2 ratios of each peptide in three replicated experiments were tested against zero by a moderated t-test from the Limma package in R/Bioconductor (Smyth, 2004), and p-values were corrected for multiple testing by the method of Benjamini-Hochberg. Peptides with a 1% FDR are considered for further analysis and are extended to the closest upstream and downstream Lys-C cleavage sites to recall the original termed 'RBDpep'. Peptides extending this set to a 10% FDR are called 'CandidateRBDpep' (Table S2).

Bioinformatic Analyses for Cardiomyocyte RBPs

Links to custom R scripts for bioinformatic analyses can be found in the R scripts availability section of the main paper.

Gene Annotation Bias

RNA binding/related terms and protein domain analyses, as well as association with Mendelian disease were essentially performed as previously described (Castello et al., 2012; Castello et al., 2013a), as were analyses of GOMF enrichment, length of protein, pI, hydrophobicity, disordered region, amino acid composition, repetitive region, low complexity region and amino acid sequence logo.

Gene Ontology and Interpro were downloaded from Ensembl release 72, 'RNA binding' annotation indicates GO:0003723 and its children terms, 'RNA-related' annotation indicates a

group of GOBP, GOCC, GOMF terms and their children terms, as well as collected Interpro terms (See R scripts).

Enrichment of GOMF categories was tested for the cardiomyocyte RBPs compared to the background of proteins identified from either the cell lysate or the cardiomyocyte RBPs (designated as WCL collectively throughout all analyses). p-values were computed by Fisher's exact test, and corrected for multiple testing by the method of Benjamini-Hochberg.

RNA-Binding Domains

RBDs listed in (Lunde et al., 2007) were considered as classic RBDs, and protein domains with experimental validation in literature were considered non-classic RBDs as described in (Castello et al., 2012).

All Pfam domains information listed in Table S1, Table S3, and RBDmap tracks were obtained through Ensembl release 72. Pfam domains information in Figure 2C/D/H were obtained through Ensembl release 84.

KEGG Pathway Analysis

To determine enrichment/depletion of KEGG pathways (Kanehisa and Goto, 2000; Kanehisa et al., 2014) in cardiomyocyte RBPs, Ensembl gene IDs were first mapped to Entrez ID using the Ensembl Biomart and then these identifiers were assigned to their respective KEGG IDs through the KEGG REST interface. WCL genes were mapped to KEGG pathways in order to establish the frequency of KEGG pathway membership in the background proteome. This resulted in a total of 238 pathways with at least one member in the WCL background. Using the same method all cardiomyocyte RBPs were then mapped to KEGG pathways, yielding a total of 200 pathways with at least one member, fully contained within the 238 WCL pathways. Pathway enrichment

was tested using Fisher's exact test, using a one-tailed test (null hypothesis: odds-ratio equals to 1). The resulting 200 p-values were corrected for multiple testing (Benjamini-Hochberg method) using a total of 238 tests, as this was the number of pathways identified in WCL. The list of cardiomyocyte RBPs was further organized into RNA-related and RNA-unrelated proteins according to their annotation as described in section 'Gene annotation bias' section of Supplemental Experimental Procedures. Pathways enriched with an FDR 10% were selected. Odds-ratios were log2 transformed, and binned as shown in Figure 1G.

STRING Analysis

Protein-protein interactions of cardiomyocyte mtRBPs were tested by STRING (version 10) (Jensen et al., 2009). The analysis was performed at a confidence level of 0.6 with seven active prediction methods, neighborhood, gene fusion, co-occurrence, co-expression, experiments, databases and textmining.

Cardiovascular-Associated RBPs

A list of cardiovascular associated (related to cardiac disease and development) proteins was downloaded from the Cardiovascular Gene Ontology Annotation Initiative website (01_2015) (www.ucl.ac.uk/cardiovasculargeneontology).

OMIM-RBPs

RBPs listed in OMIM (Ensembl BioMart, 04_ 2015) were defined as OMIM-RBPs.

mtRBPs

RBPs with GO 'mitochondrion' (GO:0005739) were defined as mtRBPs.

Metabolic Enzymes

Genes of the cardiomyocyte RBPs were mapped to 'Metabolism' events in Reactome Pathway Database (version 48) (Croft et al., 2014), the results were manually curated to remove genes whose protein products are not classified in the six enzyme commission groups (EC 1-6) from IntEnz database (Fleischmann et al., 2004), unless the gene products are subunits of enzyme complexes. This method was also applied to the HeLa (Castello et al., 2012), the HEK293 (Baltz et al., 2012) and the mESC (Kwon et al., 2013) mRNA interactomes to compare the number of RNA-binding metabolic enzymes. Metabolic pathway assignment in Table S5 is also based on Reactome Pathway Database (version 48).

Rossmann Fold (R-f) Proteins

Gene3D domain annotations for all proteins were downloaded from UniProt mouse database (release-2015_05), Figure 4C is based on cathdb release 4.0. CATH id: 3.40.50 indicates the presence of Rossmann fold. The position of Rossmann fold within a protein sequence was obtained by the sequence search tool at CATH v4.0 (Sillitoe et al., 2015).

RNA Helicases

RNA helicase families were categorized based on (Fairman-Williams et al., 2010).

RNA Modifications

Types of RNA modifications catalyzed by specific enzymes were obtained from the Modomics database (Dunin-Horkawicz et al., 2006; Machnicka et al., 2013).

Known Mitochondrial RNA-Regulating Proteins

Proteins with known involvement in the mitochondrial RNA life cycle were curated from the literature (Jourdain et al., 2013; Nicholls et al., 2013; Rackham and Filipovska, 2012; Rackham et al., 2012; Scarpulla et al., 2012; Suzuki et al., 2011).

OXPHOS Complex Subunits

Number of subunits from mitochondrial Complex I, Complex II, Complex III, Complex IV and Complex V is retrieved from HUGO Gene Nomenclature Committee (HGNC) Database (04_2015) (Gray et al., 2015).

Peptidylprolyl Isomerases (PPIases)

PPIase family members were obtained from the HUGO Gene Nomenclature Committee (HGNC) Database (04_2015) (Gray et al., 2015).

Human Orthologs

Human genes orthologous to the murine cardiomyocyte RBP genes were downloaded from Ensembl release 62. The human gene set was used for RBDpep sequence similarity analysis, analyses of cardiovascular associated annotation, Mendelian disease link and comparison to previously published mRNA interactomes. For the latter, a mouse gene was considered as possessing an ortholog in the human mRNA interactome if at least one of its orthologous human genes was contained in the human mRNA interactomes.

RS Domain Annotation

RS domain annotation for seven SRSF proteins in RBDmap was based on (Busch and Hertel, 2012; Graveley, 2000; Manley and Krainer, 2010).

Characterization of RBDpeps

Identification of Disordered Fragments

The intrinsically unstructured or disordered parts of a protein were obtained from the 'Prediction of Intrinsically Unstructured Proteins' (IUPred) database (Dosztanyi et al., 2005). Amino acids with an IUPred score higher than 0.4 were considered as disordered. An RBDpep is regarded as disordered if the average IUPred score is higher than 0.4.

Amino Acid Composition

The amino acid composition of all RBDpeps or released fragments is compared to the amino acid composition of interactome proteins. For analysis of disordered or globular RNA-binding sites, RNA bound or released proteolytic fragments overlapping with disordered or globular protein segments were compared to disordered or globular fragments in interactome proteins. Over- and under-representation of a given amino acid was tested by Fisher's exact test, and p-values were corrected for multiple testing by the method of Benjamini-Hochberg.

RBDpep Sequence Similarity Analysis

HeLa RBDmap proteins identified with Lys-C as first protease (Castello et al., 2016) were compared to HL-1 RBDmap proteins. Pairwise alignment of HL-1 RBDpeps from the overlapped proteins was performed between mouse and human homologous proteins. Similarity was calculated using the following formula:

100 * (identical positions) / (aligned positions + internal gap positions)

RBDpep Mapping to Protein Structures

Protein structures data search and download were performed at Research Collaboratory for Structural Bioinformatics (RCSB) site (www.rcsb.org) of World Wide Protein Databank (wwPDB) (Berman et al., 2003; Berman et al., 2000). The homology modeling of mouse SERCA2, RNA helicases and IDH3A structures were performed with normal modeling mode at Protein Homology/analogY Recognition Engine V 2.0 (PHYRE2) (Kelley and Sternberg, 2009). The presented structures were modeled with 100% confidence. All protein structures were viewed and analyzed with PyMOL v1.3r1.

General Molecular Biology Methods

Protein Gel Electrophoresis

Protein gel electrophoresis was performed according to standard protocols (Archer et al., 2015). DTT was added to protein samples at a final concentration of 100 mM, the samples were then boiled in 1×NuPAGE LDS sample buffer (Life Technologies) for 3 min. The proteins were loaded onto denaturing polyacrylamide gel (4-12% NuPAGE Bis-Tris gel, Life Technologies), electrophoresed at 180 V for 1 hr.

Western Blot

Gels were electroblotted onto nitrocellulose membrane (GE Healthcare Life Sciences). The membrane was blocked in 5% non-fat milk in 1×PBST (1×PBS with 0.2% Tween-20) for 30 min at room temperature, then incubated with primary antibodies on a rotating wheel o/n at 4 °C. The membrane was then washed with 1×PBST for 3 times, 5 min each at room temperature. The

membrane was incubated with secondary antibody in 5% non-fat milk in 1×PBST, for 1 hr at room temperature. Protein signals were detected using Super Signal Femto chemiluminescent reagent (Pierce), visualized on ImageQuant LAS 4000 system (GE Healthcare Life Sciences). Antibodies used were PTBP1 (SAB2101904-50UG, Sigma), ELAVL1 (ab28660, Abcam), H3 (ab1791, Abcam), H4 (ab10158, Abcam), ACTB (ab20272, Abcam), α-TUBULIN (sc-8035, Santa Cruz), CS (ab96600, Abcam), OGDH (ab137773, Abcam), IDH2 (ab131263, Abcam), ACO2 (MS793, MitoSciences), GFP (3H9, Chromotek), NSUN4 (16320-1-AP, Proteintech Group Inc.) and VDAC (4661, Cell Signaling Technology). HRP-conjugated antibodies are anti-mouse (sc-2005, Santa Cruz), anti-rat (sc-2006, Santa Cruz) and anti-rabbit (AP132P, EMD Millipore).

Silver Stain

Silver stain was performed according to standard protocol (Chevallet et al., 2006). After electrophoresis, the gel was fixed in 50% methanol (v/v), 5% acetic acid (v/v) for 30 min. The gels were washed sequentially in 50% ethanol (v/v) and 30% ethanol (v/v), 5 min for each, and then in ddH_2O for 10 min. The gel was sensitized by soaking in 0.02% sodium thiosulfate (w/v) for 1 min, followed by three washes with ddH_2O, 30 sec each. The gel was then impregnated with silver solution (6 mM silver nitrate, 0.0185% formaldehyde) for 20 min, followed by three washes with water, 30 sec each. After the final wash, developing solution (2% sodium carbonate, 0.0185% formaldehyde, 0.0004% sodium thiosulfate) was applied to the gel within 5 min of preparation. When adequate degree of staining was achieved, the gel was transferred to 10% acetic acid to stop the staining reaction.

When restain was necessary, the gel was incubated in destain solution (prepared by mixing two solutions, 30 mM potassium ferricyanide and 100 mM sodium thiosulfate, at equal volume

immediately prior to use) until gels were clear again, typically this occurred within 10 min. The gel was then washed with water to remove any yellow color. The gel was then restained as described above.

All solutions were prepared freshly, and all procedures were performed on a rocking platform at room temperature. Image of the silver stain was taken by a Nikon D5100 DSLR camera.

RNA Extraction and cDNA Synthesis

RNA from lysate or digested RNA-protein complexes was extracted with Trizol prior to random hexamer/$(dT)_{12}$VN-primed cDNA synthesis, using SuperScript III Reverse Transcriptase (Life Technologies).

Quantitative PCR

Quantitative PCR was performed using Quantifast qPCR premix (Qiagen) as per manufacturer's instructions. Serial dilutions of cDNA amplified from HL-1 total RNA were used to test the PCR efficiency, which was >90% for all primers. 10 μl reactions were generated in 384-well plates (3 technical replicates per condition) and qPCR was performed on a QuantStudio 12 K Flex (Invitrogen) (Archer et al., 2014). The following table provides PCR primer sequence information, each was used at 0.5 μM final concentration.

Gene name	***Forward primer for qPCR (5'–3')***	***Reverse primer for qPCR (5'–3')***
18s rDNA	gtaacccgttgaaccccatt	ccatccaatcggtagtagcg
Actb	gatcaagatcattgctcctcctg	agggtgtaaaacgcagctca
Gapdh	aagggctcatgaccacagtc	cagggatgatgttctgggca
Elavl1	gaagaccacatggcggaaga	cccaagctgtgtcctgctac
Mt-co1	gagaggcctttgcttcaaaaa	aggttggttcctcgaatgtg
Mt-nd3	ctagttgcattctgactcccc	atggtagacgtgcagagctt
Mt-nd2	gggcatgaggaggacttaaccaaac	tgaggttgagtagagtga
Mt-nd5	tctcacccaaaacgacatca	ttgaagaatgcgtgggtaca
Mt-atp6	aggattcccaatcgttgtagcc	ccttttggtgtgtggattagca
pT3lucpA spike-in RNA	cgccaaaagcactctgattgac	ccttgtcgtatccctggaagatg

RNA Binding Assays with eGFP-tagged Candidate RBPs

RNA 5' end-labeling assays were performed as previously described with modifications (Baltz et al., 2012; Kwon et al., 2013; Spitzer et al., 2014), using HeLa cells expressing eGFP-tagged candidate RBPs as source material.

Plasmids Construction

Plasmids that express C-terminally eGFP-tagged fusion proteins were based on an existing construct, *Nsun2-eGFP-pcDNA5/FRT/TO*, which was generated in the following way. The *Nsun2* coding sequence (CDS) was first cloned into *pEGFP-N1* (Addgene) vector through *Kpn*I/*Xma*I sites. The *Nsun2*-eGFP sequence was subsequently excised by *Kpn*I/*Not*I digest and cloned into the cognate sites of *pcDNA5/FRT/TO* (Life Technologies). To prepare the plasmids used for the

^{32}P RNA 5' end-labeling assay, the CDS from other genes were amplified from a HL-1 cDNA library using primers that introduced unique restriction sites and cloned into pGEM-T easy vector (Promega). CDS were liberated from this transit vector and replaced for the *Nsun2* portion of *Nsun2-eGFP-pcDNA5/FRT/TO* cDNA using the restriction enzyme combinations detailed in the table below. The control *eGFP-pcDNA5/FRT/TO* vector was constructed by inserting the *eGFP* CDS from *pEGFP-N1* into *pcDNA5/FRT/TO* (Life Technologies) at *Kpn*I/*Not*I sites. Details on the cloning strategy and linker sequences of the plasmids are given in the following table.

Gene name	***GenBank accession number***	***Primers for cloning (5′–3′)***	***Cut sites on cDNA ends 5′/3′***	***Cut sites on pcDNA5/FRT/TO vector 5′/3′***	***Linker amino acid sequence between fusion protein and eGFP***	***Dose of DNA ϟ***
Aco2	NM_080633.2	F*-catggtaccatctttgtcagtgcac	*Kpn*I/ *Xma*I	*Kpn*I/*Xma*I	PRDPPVAT	2
		R*-catcccggggctgctgcagctccttc				
Cs	NM_026444.3	F-catggtacccggtcccttcccgcca	*Kpn*I/ *Stu*I	*Kpn*I/*Sma*I	QGDPPVAT	2
		R-cataggcctgcttagagtccacaaac				
Idh2	NM_173011.2	F-catggtaccctcggacctcgcgtgc	*Kpn*I/ *Xma*I	*Kpn*I/*Xma*I	PRDPPVAT	0.5
		R-catcccggggctgcttgcccagagctc				
MDH2	M16229.1	F-catggtacctcctgccagtagctcc	*Kpn*I/ *Xma*I	*Kpn*I/*Xma*I	PRDPPVAT	0.5
		R-catcccggggcttcatgttcttgaca				
Pdha1	NM_008810.3	F-catggtacccgccgcgtgagtctgc	*Kpn*I/ *Xma*I	*Kpn*I/*Xma*I	PRDPPVAT	2
		R-catcccggggactgactgacttaaac				
Pum2	NM_001160222.1	F-cataagcttgcagcgcggccgcag	*Hind*III/ *Xma*I	*Hind*III/ *Xma*I	PRDPPVAT	1.2
		R-catcccggggcagcatcccatttgg				
eGFP	n/a	n/a	n/a	n/a	n/a	0.25
MLS-eGFP§	n/a	n/a	n/a	n/a	n/a	2

* F means forward primer; R means reverse primer.

ϟ means amount of DNA for Lipofectamine 2000 transfection in RNA ^{32}P 5' end-labeling assay (μg/10cm^2 culture area). Related to *Cell Transfection* section below.

§ refers to (Popow et al., 2015).

Cell Transfection

7.2×10^6 HeLa cells were seeded on each 145 cm^2 culture dish 16 hr prior to transfection. Plasmid transfections were performed with 7 ml transfection mixture containing plasmid DNA and Lipofectamine 2000 (Life Technologies) at 1:2.5 ratio (w/v). For the amount of transfected DNA and other detailed information, see the table above. Complete medium was replaced 6 hr after transfection and cells were harvested at 24 hr after transfection.

Cell Lysate Preparation

Cells were washed twice with ice-cold 1×PBS, the culture dish was then placed on ice and irradiated with 0.15 J/cm^2 UV at 254 nm. Cells were then harvested by centrifugation at 2,000 × *g* for 5 min. The cell pellets were lysed in three volumes of lysis buffer (CytoBuster protein extraction reagent, EMB Millipore), supplemented with RNaseOUT (Life Technologies) at final concentration of 100 U/ml, and complete EDTA-free protease inhibitor cocktail (Roche) at 1×, then incubated on ice for 10 min, the cell lysate was snap frozen on dry ice, thawed, and cleared by centrifugation at 10,000 × *g* for 10 min, and the supernatant was stored at -80 °C until further analysis. All centrifugations were performed at 4 °C. The typical scale of such an experiment was 3×145 cm^2 culture dishes per sample.

Preparation of Mitochondria

The procedure was performed as previously described with some modifications to accommodate the properties of cardiomyocytes (Mercer et al., 2011; Rackham et al., 2007). Cells were washed twice with ice-cold 1×PBS, the culture dish was then placed on ice and irradiated with 150 mJ/cm^2 UV at 254 nm. Cells were harvested by centrifugation at 1,000 × *g* for 5 min, and resuspended in 8 ml/145 cm^2 culture dish of homogenization buffer (250 mM sucrose, 10 mM

NaCl, 1.5 mM $MgCl_2$, 10 mM Tris•HCl, pH 7.5). Cells were then homogenized on ice, 5 times with a loose glass homogenizer, and 10 times with a tight glass homogenizer (Sigma). The suspension was centrifuged at 1,300 × *g* for 5 min and the supernatant was retained. The pellet was resuspended in sucrose buffer (5 ml/145 cm^2 culture dish; 250 mM sucrose, 10 mM Tris•HCl, pH 7.5, 1 mM EDTA), and centrifuged at 1,300 × *g* for 5 min. The two supernatants were pooled, centrifuged again at 1,300 × *g* for 5 min. The resulting supernatant was centrifuged at 12,000 × *g* for 15 min to sediment the mitochondria. The mitochondria pellet was resuspended in 1 ml sucrose buffer, and centrifuged again at 10,000 × *g* for 10 min. The supernatant was discarded. The mitochondria pellet was lysed in lysis buffer (same composition and equivalent volume as used in whole cell lysis). The mitochondria lysate was incubated on ice for 10 min, then snap frozen on dry ice, thawed, and cleared by centrifugation at 10,000 × *g* for 10 min, and the supernatant was stored at -80 °C until further analysis. All centrifugations were performed at 4 °C.

Immunoprecipitation

The cell lysate/mitochondria lysate was thawed on ice, and RNase T1 was added to the lysate at a final concentration of 1 U/µl. The reaction mixture was incubated in a water bath at 22 °C for 10 min and subsequently cooled on ice for 5 min, before addition of pre-equilibrated GFP-Trap agarose beads (Chromotek). 80 µl beads were added per ml of partial RNase T1 digested lysate. The immunoprecipitation mixture was incubated on a rotating wheel for 2 hr at 4 °C. The beads were collected by centrifugation at 2,500 × *g* for 2 min, subsequently washed once with 250 µl of high salt buffer (500 mM NaCl, 1 mM $MgCl_2$, 0.025% SDS, 0.05% NP-40, 20 mM Tris•HCl, pH 7.5), once with medium salt buffer (250 mM NaCl, 1 mM $MgCl_2$, 0.025% SDS, 0.05% NP-40, 20 mM Tris•HCl, pH 7.5), and once with low salt buffer (150 mM NaCl, 1 mM $MgCl_2$, 0.01% NP-

40, 20 mM Tris•HCl, pH 7.5). After stringent wash, the beads were resuspended in one original bead volume of low salt buffer, and RNase T1 was added to obtain a final concentration of 50 U/µl. The bead suspension was incubated in a water bath at 22 °C for 8 min, and subsequently cooled on ice for 5 min. Beads were collected by centrifugation at 2,500 × *g* for 2 min, and washed twice with wash buffer (50 mM Tris•HCl, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$). Beads were then resuspended in one original bead volume of 1×PNK buffer (New England Biolabs).

^{32}P 5' End-Labeling of RNA Segments Crosslinked to eGFP Fusion Proteins

To the bead suspension described above, γ-^{32}P-ATP (Perkin Elmer) was added to a final concentration of 0.2 µCi/µl, and T4 PNK to a final concentration of 1 U/µl.

The suspension was then incubated at 37 °C for 15 min. Thereafter, the beads were washed five times with wash buffer as above, resuspended in 30 µl elution buffer (1×NuPAGE sample buffer (Life Technologies), supplemented with SDS at final concentration 1%), and incubated at 90 °C for 5 min to elute the RNA-protein complexes. The eluate was collected by centrifugation at 2,500 × *g* for 2 min, separated on a 4-12% NuPAGE Bis-Tris gel (Life Technologies), at 180 V for 90 min, and subsequently electroblotted onto nitrocellulose membrane. The membrane was exposed to a phosphorimager screen at room temperature for 40 hr. After detecting the radioactive signal on Typhoon FLA 9000 laser scanner (GE Healthcare Life Sciences), the membrane was blocked in 5% milk in 1×PBST for 1 hr at room temperature, and eGFP fusion protein was detected by GFP antibody (3H9, Chromotek) (room temperature for 1 hr on a rocking platform). Secondary HRP conjugated anti-rat antibody incubation and protein signal detection were as described in Supplemental Experimental Procedures western blot section.

Confocal Microscopy

HeLa cells were seeded on a coverslip in 12-well cell culture dish (3.8 cm^2 culture area/well) 16 hr prior to transfection with plasmids expressing eGFP-tagged fusion proteins as described in the Supplemental Experimental Procedures *Cell transfection* section, adjusting proportionally for the smaller culture area. 24 hr after transfection, the culture media were removed, and the cells were incubated in 100 nM MitoTracker Deep Red (Life Technologies) at 37 °C for 30 min. Cells were washed three times with serum free cell culture medium at room temperature on a rocking platform, then fixed with fresh 4% paraformaldehyde at 37 °C for 15 min. The cells were then washed 3 times with 1×PBS at room temperature. Cells were mounted with VectaShield mounting medium with DAPI (VectorLabs), and imaged using Leica TCS SP5 confocal microscope.

SUPPLEMENTAL REFERENCES

Antonicka, H., and Shoubridge, E.A. (2015). Mitochondrial RNA Granules Are Centers for Posttranscriptional RNA Processing and Ribosome Biogenesis. Cell Rep. *10*, 920-932.

Archer, S.K., Shirokikh, N.E., Hallwirth, C.V., Beilharz, T.H., and Preiss, T. (2015). Probing the closed-loop model of mRNA translation in living cells. RNA Biol. *12*, 248-254.

Archer, S.K., Shirokikh, N.E., and Preiss, T. (2014). Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage. BMC Genomics *15*, 401.

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674-690.

Beckmann, B.M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., et al. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat. Commun. *6*, 10127.

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. Nat. Struct. Biol. *10*, 980.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235-242.

Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A.J. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat. Protoc. *4*, 484-494.

Busch, A., and Hertel, K.J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. WIRES RNA *3*, 1-12.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393-1406.

Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Föhr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive analysis of RNA-binding domains in human cells. Mol. Cell *In press*.

Castello, A., Fischer, B., Hentze, M.W., and Preiss, T. (2013a). RNA-binding proteins in Mendelian disease. Trends Genet. *29*, 318-327.

Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L.M., Krijgsveld, J., and Hentze, M.W. (2013b). System-wide identification of RNA-binding proteins by interactome capture. Nat. Protoc. *8*, 491-500.

Chevallet, M., Luche, S., and Rabilloud, T. (2006). Silver staining of proteins in polyacrylamide gels. Nat. Protoc. *1*, 1852-1858.

Clancy, J.L., Wei, G.H., Echner, N., Humphreys, D.T., Beilharz, T.H., and Preiss, T. (2011). mRNA isoform diversity can obscure detection of miRNA-mediated control of translation. RNA *17*, 1025-1031.

Claycomb, W.C., Lanson, N.A., Jr., Stallworth, B.S., Egeland, D.B., Delcarpio, J.B., Bahinski, A., and Izzo, N.J., Jr. (1998). HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. Proc. Natl. Acad. Sci. USA *95*, 2979-2984.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367-1372.

Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., et al. (2014). The Reactome pathway knowledgebase. Nucleic Acids Res. *42*, D472-477.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*, 3433-3434.

Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M.J., Feder, M., Grosjean, H., and Bujnicki, J.M. (2006). MODOMICS: a database of RNA modification pathways. Nucleic Acids Res. *34*, D145-149.

Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. Curr. Opin. Struct. Biol. *20*, 313-324.

Fleischmann, A., Darsow, M., Degtyarenko, K., Fleischmann, W., Boyce, S., Axelsen, K.B., Bairoch, A., Schomburg, D., Tipton, K.F., and Apweiler, R. (2004). IntEnz, the integrated relational enzyme database. Nucleic Acids Res. *32*, D434-437.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197-1211.

Gray, K.A., Yates, B., Seal, R.L., Wright, M.W., and Bruford, E.A. (2015). Genenames.org: the HGNC resources in 2015. Nucleic Acids Res. *43*, D1079-1085.

Humphreys, D.T., Hynes, C.J., Patel, H.R., Wei, G.H., Cannon, L., Fatkin, D., Suter, C.M., Clancy, J.L., and Preiss, T. (2012). Complexity of murine cardiomyocyte miRNA biogenesis, sequence variant expression and function. PLoS One *7*, e30933.

Iizuka, N., Najita, L., Franzusoff, A., and Sarnow, P. (1994). Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from Saccharomyces cerevisiae. Mol. Cell. Biol. *14*, 7322-7330.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., et al. (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. *37*, D412-416.

Jourdain, A.A., Koppen, M., Wydro, M., Rodley, C.D., Lightowlers, R.N., Chrzanowska-Lightowlers, Z.M., and Martinou, J.C. (2013). GRSF1 regulates RNA processing in mitochondrial RNA granules. Cell Metab. *17*, 399-410.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27-30.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. *42*, D199-205.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. *4*, 363-371.

Kwon, S.C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1122-1130.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479-490.

Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M., et al. (2013). MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res. *41*, D262-267.

Manley, J.L., and Krainer, A.R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). Genes Dev. *24*, 1073-1074.

Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.M., Haugen, E., Bracken, C.P., Rackham, O., Stamatoyannopoulos, J.A., et al. (2011). The human mitochondrial transcriptome. Cell *146*, 645-658.

Nicholls, T.J., Rorbach, J., and Minczuk, M. (2013). Mitochondria: mitochondrial RNA metabolism and human disease. Int. J. Biochem. Cell Biol. *45*, 845-849.

Popow, J., Alleaume, A.M., Curk, T., Schwarzl, T., Sauer, S., and Hentze, M.W. (2015). FASTKD2 is an RNA-binding protein required for mitochondrial RNA processing and translation. RNA *21*, 1873-1884.

Preiss, T., and Hentze, M.W. (1998). Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. Nature *392*, 516-520.

Rackham, O., and Filipovska, A. (2012). The role of mammalian PPR domain proteins in the regulation of mitochondrial gene expression. Biochim. Biophys. Acta *1819*, 1008-1016.

Rackham, O., Mercer, T.R., and Filipovska, A. (2012). The human mitochondrial transcriptome and the RNA-binding proteins that regulate its expression. WIRES RNA *3*, 675-695.

Rackham, O., Nichols, S.J., Leedman, P.J., Berners-Price, S.J., and Filipovska, A. (2007). A gold(I) phosphine complex selectively induces apoptosis in breast cancer cells: implications for anticancer therapeutics targeted to mitochondria. Biochem. Pharmacol. *74*, 992-1002.

Scarpulla, R.C., Vega, R.B., and Kelly, D.P. (2012). Transcriptional integration of mitochondrial biogenesis. Trends Endocrinol. Metab. *23*, 459-466.

Sillitoe, I., Lewis, T.E., Cuff, A., Das, S., Ashford, P., Dawson, N.L., Furnham, N., Laskowski, R.A., Lee, D., Lees, J.G., et al. (2015). CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res. *43*, D376-381.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, Article3.

Spitzer, J., Hafner, M., Landthaler, M., Ascano, M., Farazi, T., Wardle, G., Nusbaum, J., Khorshid, M., Burger, L., Zavolan, M., et al. (2014). PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. Methods Enzymol. *539*, 113-161.

Suzuki, T., Nagao, A., and Suzuki, T. (2011). Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu. Rev. Genet. *45*, 299-329.

Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359-362.

Yang, F., Shen, Y., Camp, D.G., 2nd, and Smith, R.D. (2012). High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. Expert Rev. Proteomics *9*, 129-134.